DOCUMENT RESUME

ED 038 646 24 AL 002 394

| | |
|---|---|
| AUTHOR | Ting, Aichen; And Others |
| TITLE | Phonetic Transcription: A Study of Transcriber Variation. |
| INSTITUTION | Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning. |
| SPONS AGENCY | Office of Education (DHEW), Washington, D.C. Bureau of Research. |
| REPORT NO | TP-122 |
| BUREAU NO | BR-5-0216 |
| PUB DATE | Mar 70 |
| CONTRACT | OEC-5-10-154 |
| NOTE | 22p.; Report from the Project on Language Concepts and Cognitive Skills Related to the Acquisition of Literacy |
| EDRS PRICE | EDRS Price MF-$0.25 HC-$1.20 |
| DESCRIPTORS | *Analysis of Variance, Articulation (Speech), Child Language, *Phonemes, *Phonetic Transcription, Pronunciation, *Reliability |

ABSTRACT

This study examined variation in transcriber disagreement as a function of transcriber's linguistic background, the transcription task, and the nature of judgment involved. Three linguistics students trained in phonetic transcription listened to the same tapes of Midwestern kindergarteners pronouncing lists of common words. Transcription task varied with order of listening: the first transcriber listened for errors of articulation and transcribed them in broad phonetic notation. The other two transcribers served as checkers of the first transcription. Disagreements between the first transcriber and first checker varied as a function of task and judgment but not as a function of the individuals' linguistic backgrounds. The first transcriber adopted the stricter criterion of correct pronunciation; the first checker appeared to expect an error in each word heard, with a consequently greater disagreement rate for sounds judged correct by the first transcriber when they appeared in words judged correct. The judgment of whether or not a sound in a word was mispronounced produced, at most, only half as many disagreements as the selection of a particular transcription for a sound thought to be in error by both transcribers. A correction procedure for transcription seems necessary for any study of articulation assessing the nature of the errors. (Author/AMM)

BR. 5-0216
PA-24

TECHNICAL REPORT NO. 122

# PHONETIC TRANSCRIPTION: A STUDY OF TRANSCRIBER VARIATION

REPORT FROM THE PROJECT ON LANGUAGE CONCEPTS AND COGNITIVE SKILLS RELATED TO THE ACQUISITION OF LITERACY

ED 038646

WISCONSIN RESEARCH AND DEVELOPMENT

CENTER FOR COGNITIVE LEARNING

THE UNIVERSITY OF WISCONSIN
Madison, Wisconsin

U.S. Office of Education
Center No. C-03
Contract OE 5-10-154

UNIVERSITAS WISCONSINENSIS
NUMEN LUMEN
SIGILLUM

RESEARCH
COGNITIVE LEARNING
DEVELOPMENT

ED038646

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Technical Report No. 122

# PHONETIC TRANSCRIPTION: A STUDY OF TRANSCRIBER VARIATION

By Aichen Ting, Richard L. Venezky,
Robin S. Chapman, and Robert C. Calfee

Report from the Project on Language Concepts and
Cognitive Skills Related to the Acquisition of Literacy

Robert C. Calfee and Richard L. Venezky,
Principal Investigators

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

March 1970

AL 002 394

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

## NATIONAL EVALUATION COMMITTEE

**Samuel Brownell**
Professor of Urban Education
Graduate School
Yale University

**Launor F. Carter**
Senior Vice President on Technology and Development
System Development Corporation

**Francis S. Chase**
Professor
Department of Education
University of Chicago

**Henry Chauncey**
President
Educational Testing Service

**Martin Deutsch**
Director, Institute for Developmental Studies
New York Medical College

**Jack Edling**
Director, Teaching Research Division
Oregon State System of Higher Education

**Elizabeth Koontz**
Wage and Labor Standards Administration, U.S. Department of Labor, Washington

**Roderick McPhee**
President
Punahou School, Honolulu

**G. Wesley Sowards**
Director, Elementary Education
Florida State University

**Patrick Suppes**
Professor
Department of Mathematics
Stanford University

***Benton J. Underwood**
Professor
Department of Psychology
Northwestern University

## RESEARCH AND DEVELOPMENT CENTER POLICY REVIEW BOARD

**Leonard Berkowitz**
Chairman
Department of Psychology

**Archie A. Buchmiller**
Deputy State Superintendent
Department of Public Instruction

**Robert E. Grinder**
Chairman
Department of Educational Psychology

**Russell J. Hosler**
Professor, Curriculum and Instruction

**Clauston Jenkins**
Assistant Director
Coordinating Committee for Higher Education

**Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational Psychology

**Stephen C. Kleene**
Dean, College of Letters and Science

**Donald J. McCarty**
Dean
School of Education

**Ira Sharkansky**
Associate Professor of Political Science

**B. Robert Tabachnick**
Chairman, Department of Curriculum and Instruction

**Henry C. Weinlick**
Executive Secretary
Wisconsin Education Association

**M. Crawford Young**
Associate Dean
The Graduate School

## EXECUTIVE COMMITTEE

**Edgar F. Borgatta**
Brittingham Professor of Sociology

**Anne E. Buchanan**
Project Specialist
R & D Center

**Robin S. Chapman**
Research Associate
R & D Center

**Robert E. Davidson**
Assistant Professor, Educational Psychology

**Frank H. Farley**
Associate Professor, Educational Psychology

**Russell J. Hosler**
Professor of Curriculum and Instruction and of Business

***Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational Psychology

**Wayne Otto**
Professor of Curriculum and Instruction (Reading)

**Robert G. Petzold**
Associate Dean of the School of Education
Professor of Curriculum and Instruction and of Music

## FACULTY OF PRINCIPAL INVESTIGATORS

**Vernon L. Allen**
Professor of Psychology

**Ted Czajkowski**
Assistant Professor of Curriculum and Instruction

**Robert E. Davidson**
Assistant Professor of Educational Psychology

**Gary A. Davis**
Associate Professor of Educational Psychology

**M. Vere DeVault**
Professor of Curriculum and Instruction (Mathematics)

**Frank H. Farley**
Associate Professor of Educational Psychology

**Lester S. Golub**
Lecturer in Curriculum and Instruction and in English

**John G. Harvey**
Associate Professor of Mathematics and of Curriculum and Instruction

**Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational Psychology

**Donald Lange**
Assistant Professor of Curriculum and Instruction

**James Moser**
Assistant Professor of Mathematics Education; Visiting Scholar

**Wayne Otto**
Professor of Curriculum and Instruction (Reading)

**Milton O. Pella**
Professor of Curriculum and Instruction (Science)

**Thomas A. Romberg**
Associate Director, R & D Center
Professor of Mathematics and of Curriculum and Instruction

**B. Robert Tabachnick**
Chairman, Department of Curriculum and Instruction

**Richard L. Venezky**
Assistant Professor of English and of Computer Sciences

**Alan Voelker**
Assistant Professor of Curriculum and Instruction

**Larry Wilder**
Assistant Professor of Curriculum and Instruction

**Peter Wolff**
Assistant Professor of Educational Psychology

## MANAGEMENT COUNCIL

**Herbert J. Klausmeier**
Director, R & D Center
V.A.C. Henmon Professor of Educational Psychology

**Mary R. Quilling**
Director
Technical Development Program

**Thomas A. Romberg**
Associate Director

**James Walter**
Director
Dissemination Program

**Dan G. Woolpert**
Director
Operations and Business

* COMMITTEE CHAIRMAN

## STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from the Language Concepts and Cognitive Skills Related to the Acquisition of Literacy Project in Program 1. General objectives of the Program are to generate new knowledge about concept learning and cognitive skills, to synthesize existing knowledge, and to develop educational materials suggested by the prior activities. Contributing to these Program objectives, this project's basic goal is to determine the processes by which children aged four to seven learn to read, examining the development of related cognitive and language skills, and to identify the specific reasons why many children fail to learn to read. Later studies will be conducted to find experimental techniques and tests for optimizing the acquisition of skills needed for learning to read. By-products of this research program include methodological innovations in testing paradigms and measurement procedures; the present study is an example.

## CONTENTS


| | | Page |
|---|---|---|
| | List of Tables | vii |
| | Abstract | ix |
| I | Introduction | 1 |
| | Studies of Reliability | 1 |
| | Factors Which Influence Transcription | 2 |
| II | Method and Results | 5 |
| III | Discussion | 9 |
| | References | 11 |
| | Appendix | 13 |

## LIST OF TABLES


| Table | | Page |
|---|---|---|
| 1 | Transcriber Disagreement by Phoneme for Words in Which T1 Found an Error | 6 |
| 2 | Transcriber Disagreement on Phonemes Transcribed as Errors by Both T1 and C1 | 6 |
| 3 | Transcriber Disagreement on Phonemes Found Incorrect by T1 and Correct by C1 | 7 |
| 4 | Transcriber Disagreement on Phonemes Found Correct by T1 in Words for Which T1 Transcribed an Error | 7 |
| 5 | Transcriber Disagreement on Phonemes in Sample of Words Treated as Correct by T1 | 7 |
| 6 | Proportion of Second Checker's Choices of Transcription for T1-C1 Disagreements in Words in Which T1 Found an Error | 8 |

## ABSTRACT

This study examined variation in transcriber disagreement as a function of transcriber's linguistic background, the transcription task, and the nature of judgment involved. Three linguistics students trained in phonetic transcription, one a non-native speaker of English, listened to the same tapes of Midwestern Kindergarteners pronouncing lists of common words. Transcription task varied with order of listening; the first transcriber listened for errors of articulation and transcribed them in broad phonetic notation. The other two transcribers served as checkers of the first transcription. The first checker independently transcribed errors for the words in which the first transcriber had found errors [an undifferentiated sample of words found correct was included in the first checker's set of items]. The second checker listened to words for which the two transcriptions differed and selected one of two transcriptions as correct, or added her own. Five protocols for each of the six possible combinations of first transcriber, first checker, and second checker were selected and examined for disagreement.

Disagreements between the first transcriber and first checker varied as a function of task and judgment but not as a function of the individuals' linguistic backgrounds. The first transcriber adopted the stricter criterion of correct pronunciation; the first checker appeared to expect an error in each word heard, with a consequently greater disagreement rate for sounds judged correct by the first transcriber when they appeared in words judged correct. The judgment of whether or not a sound in a word was mispronounced produced, at most, only half as many disagreements as the selection of a particular transcription for a sound thought to be in error by both transcribers. In the latter disagreements, the first checker's transcription was selected as correct by the second checker 70% of the time, irrespective of the identity of the checkers. On the basis of these findings, it is argued that a correction procedure for transcription is necessary for any study of articulation assessing the nature of the errors.

# I
# INTRODUCTION

Analysis of speech, the sine qua non for accurate studies of language habits, begins, for practical reasons, with the reduction of a complex acoustical stream to a visual record. For analysis of certain acoustical features of speech and for limited and not very accurate recognition of segmental units, electronic devices can be employed. But on the majority of those occasions when speech is examined, the human listener is enlisted for the transcription stage. The ability of listeners with linguistic training to transcribe speech accurately is the subject of the present investigation.

The purpose of this study is to examine transcriber ability by determining the degree of disagreement between pairs of trained linguists in the transcription of data gathered in a developmental study of articulation. Transcriber disagreements are also examined for systematic variation as a function of transcriber characteristics and transcription task.

## STUDIES OF RELIABILITY

Linguists, while aware of the possibilities of observer bias in recording speech, have generally attempted to optimize transcriber accuracy through training, without special concerns for the theoretical degree of reliability possible or the actual degree of reliability obtained. In some studies, variation among transcribers [though not inconsistancy in a single transcriber's work] was eliminated through the use of a single, well-trained linguist who did all of the transcribing. Such was the procedure adopted in the late 19th Century for a dialect atlas of French, when Edmont Edmond was sent by bicycle through France and adjoining areas to interview some 600 informants.[1] Most large dialect studies, however, have enlisted teams of fieldworkers, depending upon past experience and training to produce agreement among them. The first group of fieldworkers for the Linguistic Atlas of the United States, mostly linguists by trade, undertook 6 weeks of extensive training in fieldwork in the summer of 1930 before beginning to transcribe the nuances of New England speech. Following the completion of the New England fieldwork, each of the two directors of the project rank-ordered the fieldworkers (eight, including themselves) on a number of specific skills, including:

1. Minuteness in phonetic recording;
2. Freedom from systematization according to the fieldworker's phonemic system;
3. Freedom from systematization according to the informant's phonemic system;
4. Avoidance of over-transcription; and
5. Accuracy in recording quantity and stress.

There was considerable variation in the orderings for several of these categories, indicating some independence among these skills. No attempt was made, however, to assess overall reliability by comparing transcriptions of a common source.

In psychologically oriented studies where speech transcriptions are done, reliability checks range from none to extensive independent studies. Templin (1957), in a study of various language skills in 480 children, dispensed completely with reliability or consistency checks:

---

[1] See J. Gillieron and E. Edmond, Atlas linguistique de la France. Paris, 1902-10.

Since the author has had a substantial amount of training and experience in the use of phonetics and a good deal of experience in the testing of speech of young children, and because of the difficulties in introducing a third person into the test situation with preschool children, no reliability check was made. Repeated tests in the same child would not provide several judgments of the same utterance.

Henderson (1935), in a study of articulation in normal institutionalized children, measured consistency, i.e., within-transcriber reliability, by comparing two transcriptions which she made of the same phonographic recording. She found that 98.2% of her own transcriptions were identical. While this technique does measure intra-judge consistency, it does not evaluate accuracy. Consistently aberrant broad transcriptions could rank higher on this measure than accurate narrow transcriptions varying in some minute detail of vowel length or degree of voicing.

In a separate study, Henderson (1937) had three judges record five consonants which occurred in real words pronounced for them by a young child. Two instances of each consonant in initial, medial, and final positions were heard. When judges' responses were limited to "correct" or "incorrect," the three judges agreed 80% of the time; when a phonetic representation of the sound was required, however, three-way agreements dropped to 72%. When the judges performed the same tasks, but with a 5-year old child heard from a loud-speaker rather than seen, the agreement figures dropped to 69% and 60%, respectively. Nevertheless, these tasks are considerably easier than those the transcriber faced in the earlier Henderson study, which required transcription of whole utterances, rather than a single consonant from each.

Irwin and Curry (1941) and Irwin and Chen (1941) tested agreement between pairs of transcribers who were recording crying vocalizations of infants under 10 days of age. While the agreements were relatively high (85% for vowels and 94% for consonants), the task is quite distinct from recording real speech. The chief difference is that the repertoire of crying sounds is small. Furthermore, while both sets of authors made transcriptions of specific phonemes such as /æ/ and /ɪ/, the existence of such entities in the speech of 1- to 9-day-old children is doubtful. Phonemes depend for their existence upon a system of contrasts which function to separate meanings. It is highly unlikely that any of the sounds made by a 1- to 9-day-old child meet this criterion. While crying may contain vowel-like sounds, it does not contain vowels in the same sense that the speech of a normal 5-year old does. The acoustic patterns are distinct. At best, one can interpret the phonemes presented in these studies as indicators of the allophones which most closely approximate the sounds which the children emitted.

## FACTORS WHICH INFLUENCE TRANSCRIPTION

The factors which are most important for agreement between transcribers appear to be hearing, training in phonetics, familiarity with the speech to be transcribed, and degree of detail required. While detection of sound pressure and frequency differences are important for hearing any sound, little is known about hearing abilities specifically related to speech reception, aside from lateralization, which is peculiar to dichotic listening, e.g., Kimura (1961).

Training in phonetics, like training in art, is a necessary prerequisite for detecting details in the material. Just as the untrained eye will be unaware of the nuances of brush stroke, perspective, and division of space, so will the untrained listener be unaware of the finer shades of aspiration, nasalization, and devoicing, and how to represent them in writing. Along with training in general phonetics must go training or familiarization in the language or dialect being transcribed.

A fieldworker will often detect a difference between two sounds, yet not know, on first listening, the exact nature of the difference. Through repeated listening, comparisons to similar sounds, and attempts at imitation, he will generally uncover the phonetic basis for the distinction. From that point on he will be attuned for such forms. An untrained fieldworker can often perform reasonably well when transcribing a familiar dialect in broad phonetic or phonemic notation, but will begin to trip over unfamiliar speech patterns or the need to mark narrow details.

In testing inter-judge reliability, therefore, it is important to note transcriber, speech stimulus, and task characteristics. The three transcribers compared in the present study were all trained in phonetic transcription, but one was a non-native speaker of English; all were living in the dialect area of the speech transcribed. The materials to be transcribed were tape recordings of young children repeating common words; the transcribers could listen to each

recording as often as they chose. The first transcriber listened for errors of articulation and transcribed in broad phonetic notation (International Phonetic Alphabet) the substituted or inserted sounds, or noted deletions. The second transcriber independently transcribed errors for the subset of words in which the first transcriber had found errors, plus an undifferentiated random selection of words for which no errors had been transcribed. The third decided between differing transcriptions made by the first and second transcribers. Transcribers could disagree in judgment at two levels: they could disagree about whether a sound was in error; or agree that the sound was in error but disagree over its transcription. The data were analyzed for differences in agreement as a function of transcriber identity, task, and judgment.

## II
## METHOD AND RESULTS

A comparison was made of transcriptions of three transcribers who listened to the same tape-recorded materials and recorded errors of articulation in broad IPA. The transcribers were:

A: A female undergraduate in linguistics, who had spent most of her life in New Jersey aside from 4 years of high school in Florida. She had studied both French and Russian for 5 years, Chinese for 1 year, German for half a year, and had had approximately 1 year of experience in phonetic transcription [aside from course work in linguistics].

B: A female graduate student in linguistics, who was raised in Northern Virginia, but attended college in Massachusetts. She studied German for 7 years including 1 year in Germany, French for 5 years, Chinese for 2 years, Spanish, Latin, and Greek for 1 year, and Hindi for half a year. She had had approximately 2 years of experience in phonetic transcription [aside from course work in linguistics].

C: A female graduate student in linguistics, born in Peking, China, who speaks two dialects of Chinese—Pekingese and Cantonese. She studied British English as a foreign language in high school in Hong Kong, and American English in college in Taiwan. She then spent 2 years in New York City and 2 years in Providence, R. I. In addition to English, she studied French for 1 year and had had approximately 2 years of experience in phonetic transcription [aside from course work in linguistics].

The materials to be transcribed were tapes of Midwestern Kindergarten children repeating standardized lists of common words. The articulation lists included two random orderings of each of two different 48 word lists. Most were high-frequency words, chosen to test each vowel of English in at least two environments, single consonants and consonant clusters in initial and final position, and three-item consonant clusters in initial position. (The two lists of words are included in the Appendix.)

Ss' pronunciations were recorded on a Uher 5000 tape recorder with a Shure lavaliere microphone at 3 3/4 ips. Tapes were listened to on a Uher 5000 tape recorder. The transcriber could listen to an item as many times as she needed, and had data sheets giving the spelling of each stimulus word, but not a phonetic representation. Transcribers recorded any errors detected in broad IPA; appropriate pronunciations of words were not transcribed. Transcription symbols were limited to the phonemes of English plus glottal stops, bilabial fricatives, and aspiration diacritics. No judgments of vowel lengthening or shortening, final release, weakly articulated consonants, or other narrow phonetic details were required.

One of the three transcribers listened to the S repeat the whole word list, transcribing errors in broad IPA; the other two served as checkers. All words in which the first transcriber found an error were marked for future checking. Furthermore, 10% of a subject's responses were randomly selected from words for which no error had been transcribed by the first transcriber and added to the items to be checked; these were not differentiated from error words. The first checker covered up the original transcription, listened to the items to be checked, and transcribed the errors or indicated that no error was present. If the first checker disagreed with the original transcriber, then the second checker listened to the item and indicated which of the two transcriptions she agreed with. She could also add a third transcription if she could not agree with either of the first two.

The possible patterns of disagreement arising in the course of checking are as follows: (T1 stands for the first transcriber; C1, the first checker; C2, the second checker.)

Case 1. C1 agrees with T1. No disagreements.

Case 2. C1 disagrees with T1.

2a. C2 agrees with T1. Two-person disagreement.
2b. C2 agrees with C1. Two-person disagreement.
2c. C2 adds a new transcription. Three-person disagreement.

Since each of the three transcribers could serve as first transcriber, first checker, or second checker, five transcription sheets with checks were randomly selected, from Kindergarten subjects, for each of the following combinations of transcriber, first checker, and second checker (A, B, and C refer to the transcribers). Thus, 30 transcription records comprise the sample for this study.

| Combination | T1 | C1 | C2 |
|---|---|---|---|
| 1 | A | B | C |
| 2 | A | C | B |
| 3 | B | A | C |
| 4 | B | C | A |
| 5 | C | A | B |
| 6 | C | B | A |

## RESULTS

A tabulation of two-person transcriber disagreements by phoneme for the words in which T1 found an error is given in Table 1. The count for phonemes listened to by both T1 and C1 was based on the expected number of phonemes for correct responses to each word.

In Tables 2-4, these disagreements for words in which T1 transcribed an error are broken down as follows: a) T1 and C1 transcribed a different error for the same expected sound (Table 2); b) C1 found the sound T1 transcribed as an error to be correct (Table 3); C1 found an error in one of the sounds considered correct by T1 (Table 4).

Disagreements between T1 and C1 could also arise when C1 found a phoneme error in the sample of words for which T1 had recorded no error. In Table 5 are presented the phoneme disagreements for words indicated to be correct by T1 and checked by C1. Again, the count for phonemes listened to by both T1 and C1 was based on the expected number of phonemes for correct responses to each word.

The phoneme disagreement rates reported in Tables 2-5 show little variation as a result of transcriber pairing but marked variation between tables. Most disagreements arose over sounds transcribed as errors by T1; C1 disagreed with T1's transcription 169 out of 499 times, or 33.6% of the time. The majority (70%) of these disagreements occurred because C1 transcribed a different error than T1 for the sound in question

Table 1

Transcriber Disagreement by Phoneme for Words in which T1 Found an Error[a]

| T1 - C1 | No. of Phonemes Listened to by T1 - C1 | No. of Phoneme Disagreements | Percentage |
|---|---|---|---|
| A - B | 252 | 32 | 12.7% |
| A - C | 414 | 40 | 9.6% |
| B - A | 181 | 16 | 8.8% |
| B - C | 208 | 24 | 11.5% |
| C - A | 313 | 37 | 11.8% |
| C - B | 311 | 28 | 9.0% |
| Total | 1679 | 177 | 10.5% |

[a]Five S protocols are represented in each T1-C1 category.

Table 2

Transcriber Disagreement on Phonemes Transcribed as Errors by both T1 and C1[a]

| T1 - C1 | No. of Phoneme Errors Transcribed by T1 | No. of Phoneme Disagreements | Percentage |
|---|---|---|---|
| A - B | 72 | 18 | 25.0% |
| A - C | 131 | 25 | 19.0% |
| B - A | 57 | 10 | 17.5% |
| B - C | 63 | 19 | 30.1% |
| C - A | 91 | 25 | 27.4% |
| C - B | 85 | 21 | 24.7% |
| Total | 499 | 118 | 23.6% |

[a]Five S protocols are represented in each T1-C1 category.

Table 3

Transcriber Disagreement on Phonemes Found Incorrect by T1 and Correct by C1[a]

| T1 - C1 | No. of Phoneme Errors Transcribed by T1 | No. of Phoneme Disagree-ments | Percent-age |
|---|---|---|---|
| A - B | 72 | 13 | 18.0% |
| A - C | 131 | 12 | 9.0% |
| B - A | 57 | 5 | 8.0% |
| B - C | 63 | 5 | 8.0% |
| C - A | 91 | 9 | 9.8% |
| C - B | 85 | 7 | 8.0% |
| Total | 499 | 51 | 10.2% |

[a]Five S protocols are represented in each T1-C1 category.

Table 4

Transcriber Disagreement on Phonemes Found Correct by T1 in Words for which T1 Transcribed an Error[a]

| T1 - C1 | No. of Phonemes Listened to by T1-C1 | No. of Phoneme Disagree-ments | Percentage |
|---|---|---|---|
| A - B | 180 | 1 | 0.5% |
| A - C | 283 | 3 | 0.1% |
| B - A | 124 | 1 | 0.8% |
| B - C | 145 | 0 | 0.0% |
| C - A | 222 | 3 | 0.1% |
| C - B | 226 | 0 | 0.0% |
| Total | 1180 | 8 | 0.6% |

[a]Five S protocols are represented in each T1-C1 category.

(Table 2). C1 disagreed with T1's judgment that the sound was in error 51 out of 499 times, or 10.2% (Table 3). In contrast for those words in which T1 found an error, C1 recorded an error for the sounds of the word judged correct by T1 only .6% of the time (Table 4). The disagreement rate rises when words which T1 found correct are considered; here, C1 disagreed with T1's judgment of no phoneme error 4.6% of the time (Table 5).

Table 5

Transcriber Disagreement on Phonemes in Sample of Words Treated as Correct By T1[a]

| T1 - C1 | No. of Phonemes Listened to by T1-C1 | No. of Phoneme Disagree-ments | Percent-age |
|---|---|---|---|
| A - B | 117 | 9 | 7.6% |
| A - C | 78 | 2 | 2.5% |
| B - A | 76 | 2 | 2.6% |
| B - C | 87 | 4 | 4.5% |
| C - A | 82 | 0 | 0.0% |
| C - B | 119 | 9 | 7.5% |
| Total | 559 | 26 | 4.6% |

[a]Five S protocols are represented in each T1-C1 category.

The phoneme disagreements between T1 and C1 were resolved by a second checker (C2), who chose between the two transcriptions or added one of her own, which was taken as the authorative transcription. Among the 177 phoneme disagreements of T1 and C1 arising for the words in which T1 had recorded an error, C2 agreed with T1 44 times, or 25%, irrespective of the identity of C2. C2 agreed with C1 124 times, or 70%, again irrespective of the identity of C2. C2 added a new transcription only 5% of the time, irrespective of the identity of T1, C1, and C2, The proportion of times C2 agreed with each transcriber serving as T1 or C1 is given in Table 6. Most of the variation is accounted for by the task (T1 or C1) of the transcriber, rather than her identity; a particular C2 always agreed more often with another transcriber when she served as C1. In the case of A serving as C2, however, there is evidence of a tendency to agree with B more often than C.

Table 6

Proportion of Second Checker's Choices of Transcription for T1-C1 Disagreements in Words in Which T1 Found an Error[a]

| Second Checker Agreed With: | Second Checker | | |
|---|---|---|---|
| | A | B | C |
| A | (0.05)[b] | .43 | .46 |
| as T1 only | --- | .30 | .28 |
| as C1 only | --- | .70 | .88 |
| B | .68 | (.05)[b] | .49 |
| as T1 only | .48 | --- | .07 |
| as C1 only | .85 | --- | .67 |
| C | .27 | .52 | (.05)[b] |
| as T1 only | .10 | .20 | --- |
| as C1 only | .48 | .56 | --- |
| T1 | .22 | .28 | .26 |
| C1 | .73 | .67 | .69 |
| (C2) | (.05) | (.05) | (.05) |

[a]Five S protocols are represented in each T1-C1 category.

[b]Five percent of the time, the second checker selected her own transcription as correct, rather than T1's or C1's.

## III
## DISCUSSION

In the present study of transcriber variation, transcribers did not see the children pronouncing words but were allowed to listen to each word as often as they needed. Henderson's study (1937) indicates that loss of visual information reduces transcriber agreement when only one opportunity to listen to the word is given. When more opportunities are given, one would expect an increase in agreement, but there are no data on this point. A relatively broad IPA transcription was required; more detailed transcription would presumably increase disagreement rate. Under these conditions of transcription, the most striking finding is the difference in disagreement rate as a function of the transcriber's task and judgment required, rather than the background of individual transcribers. For sounds found correct by the first transcriber, two-person disagreements ranged from .6%, when the first transcriber had found an error elsewhere in the word, to 4.6%, when she had found no error in the word. For sound transcribed as errors by the first transcriber two-person disagreements rose to 33.8%; 23.6% in which the first checker provided a different transcription of the error and 10.6% in which she thought the sound correct, rather than in error. These patterns of disagreement were true for any pair of transcribers.

Recall that the first checker's primary function was to check the first transcriber's transcription of errors; that is, she expected to find errors in most of the words she listened to. The difference in disagreement for sounds found correct by the first transcriber (.6% when T1 had found an error elsewhere in the word; 4.6% when not) appears to stem from this expectation: the first checker, in effect, listened especially hard for errors in words until she found one—thus perhaps increasing the probability of recording one for non-error words but decreasing it for sounds in words for which she had already located an error.

That the first checker considered 10.2% of the errors identified by T1 to be correct pronunciations, rather than errors, indicates that the individual adopted a sterner criterion for correct pronunciation as first transcriber than as first checker, reflecting the attitude that the primary burden of transcription was on the first transcriber. These differences in disagreement over whether a sound is correct or incorrect, then, appear to arise from the task assigned the individual for a given protocol. Differences among the three individuals when they all have the same task assignment (e.g., first checker) are minimal; the different disagreement rates related to task, however, are marked even within an individual.

Variation in disagreement can not only be traced to task differences and associated expectations, but also to differences in the nature of the judgment required. A transcriber must determine a) whether a given sound in a word is mispronounced and b) if it is, decide exactly what the error is and transcribe it. The first judgment is considerably easier than the second, according to both Henderson's data (1937) and the present study. Maximum disagreement over whether a sound was correctly pronounced or not was 10.2%, when the first transcriber judged the pronunciation incorrect. Given that the transcriber and checker agreed that a sound was in error, however, they disagreed as to the exact transcription of the sound 23.6% of the time.

The high rate of disagreement over the exact transcription for an error may simply reflect the difficulty in choosing a transcription for a sound that may not even be an English allophone, or it may additionally reflect inaccuracies associated with the task of first transcriber. If the first checker was listening with the expectation of hearing an error, then presumably more of her attention could be given to the exact nature of the mispronunciation and less to the correct-incorrect decision required

of the first transcriber. Our data indicate that all three transcribers regard their colleagues' transcriptions as more accurate when the colleagues have served as first checker rather than first transcriber—70% of the time the first checker's transcription is chosen.

In summary, the data indicate that transcriber agreement is a function of the task assigned and the specificity of judgment required but not, in this case, of the individual's language background. The first transcriber used a stricter criterion for correct pronunciation than the checkers and noted errors at the cost of noting their exact nature. The data suggest that a two-part transcription procedure would be more reliable for the transcription of articulatory errors, when the material can be listened to more than once: a first pass through the material in which the transcriber's sole task is to mark the places in which he hears mispronunciations, and a second pass through the material to transcribe those errors. Such a procedure would allow the first transcriber to devote his attention to the exact nature of the mispronunciation. Whether such a transcription procedure is adopted or not, it should be noted that fully 25% of the first transcriber's error transcriptions were ultimately changed. The fact that each individual was agreed with more often as first checker than as first transcriber indicates that these changes can be regarded as corrections. The magnitude of their number would seem to make such correction procedures mandatory for studies which focus on the nature, as opposed to the number, of articulatory errors in children's speech.

## REFERENCES


Henderson, F. A study of the articulation of consonants by normal institutionalized children. Unpublished Ph.D. thesis, University of Wisconsin, 1935.

Henderson, F. Objectivity and constancy of judgment in articulation testing. Journal of Educational Research, 1937, 31, 348-56.

Irwin, O. C., & Chen, H. P. A reliability study of speech sounds observed in the crying of newborn infants. Child Development, 1941, 12, 351-68.

Irwin, O. C., & Curry, T. Vowel elements in the crying vocalization of infants under ten days of age. Child Development, 1941, 12, 99-109.

Kimura, D. Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology, 1961, 15, 166-71.

Templin, M. C. Certain language skills in children. Institute of Child Welfare, Monograph No. 26, 1957. Minneapolis: The University of Minnesota Press.


GPO 816-845-3

Appendix

Articulation Test 67-FA
Order I
Condition _____ Voice _____

No. _____ Name _____ Age ____

Tape No. _____

School _____ Grade _____

Side _____

Date Transcribed _____

Footage: From _____ To _____

Transcriber _____

| LIST A | ERRORS | ERROR CHECK | NOTES |
|---|---|---|---|
| 1. smoke | | | |
| 2. frog | | | |
| 3. string | | | |
| 4. playing | | | |
| 5. health | | | |
| 6. clocks | | | |
| 7. beige | | | |
| 8. yawn | | | |
| 9. cold | | | |
| 10. noise | | | |
| 11. girl | | | |
| 12. quack | | | |
| 13. glass | | | |
| 14. vice | | | |
| 15. tenth | | | |
| 16. salt | | | |

Articulation Test 67-FA (continued)

| LIST A | ERRORS | ERROR CHECK | NOTES |
|---|---|---|---|
| 17. zoo | | | |
| 18. foot | | | |
| 19. bloom | | | |
| 20. throw | | | |
| 21. dish | | | |
| 22. them | | | |
| 23. scarf | | | |
| 24. grass | | | |
| 25. twine | | | |
| 26. judge | | | |
| 27. length | | | |
| 28. red | | | |
| 29. birth | | | |
| 30. spray | | | |
| 31. church | | | |
| 32. coins | | | |
| 33. thumb | | | |
| 34. wind | | | |
| 35. star | | | |
| 36. tree | | | |
| 37. flowers | | | |
| 38. mouth | | | |
| 39. proud | | | |
| 40. crib | | | |
| 41. sleep | | | |
| 42. drink | | | |
| 43. splashing | | | |
| 44. scratch | | | |

Articulation Test 67-FA (continued)

| LIST A | ERRORS | ERROR CHECK | NOTES |
|---|---|---|---|
| 45. sheep | | | |
| 46. breathe | | | |
| 47. swimming | | | |
| 48. pull | | | |

Articulation Test 67-FB
Order I
Condition _____ Voice _____

No. _____ Name _____ Age __

Tape No. _____

School _____ Grade ____

Side _____

Date Transcribed _____

Footage: From _____ To _____

Transcriber _____

| LIST B | ERRORS | ERROR CHECK | NOTES |
|---|---|---|---|
| 1. through | | | |
| 2. friend | | | |
| 3. crown | | | |
| 4. flash | | | |
| 5. spring | | | |
| 6. tooth | | | |
| 7. young | | | |
| 8. grief | | | |
| 9. price | | | |
| 10. third | | | |
| 11. zoom | | | |
| 12. cage | | | |
| 13. school | | | |
| 14. wharf | | | |
| 15. glad | | | |
| 16. good | | | |
| 17. fog | | | |
| 18. month | | | |
| 19. slow | | | |
| 20. pink | | | |
| 21. that | | | |
| 22. broil | | | |
| 23. stop | | | |

Articulation Test 67-FB (continued)

| LIST B | ERRORS | ERROR CHECK | NOTES |
|---|---|---|---|
| 24. voice | | | |
| 25. close | | | |
| 26. scrub | | | |
| 27. reach | | | |
| 28. shirt | | | |
| 29. swinging | | | |
| 30. split | | | |
| 31. just | | | |
| 32. house | | | |
| 33. lens | | | |
| 34. wealth | | | |
| 35. child | | | |
| 36. knob | | | |
| 37. smile | | | |
| 38. bathe | | | |
| 39. queen | | | |
| 40. truck | | | |
| 41. plum | | | |
| 42. strength | | | |
| 43. twins | | | |
| 44. door | | | |
| 45. blouse | | | |
| 46. six | | | |
| 47. drive | | | |
| 48. built | | | |

GPO 816-845-2